Automatic Classification of Cetacean Vocalizations Using an Aural Classifier

Paul C. Hines & Carolyn M. Binder
Defence R&D Canada–Atlantic
PO Box 1012 Dartmouth
NS, Canada, B2Y3Z7
phone: (902) 426-3100 fax: (902) 426-9654 email: paul.hines@drdc-rddc.gc.ca

Award Number: N00014-12-1-0139
http://www.drdc-rddc.gc.ca/

LONG-TERM GOALS

To develop a robust automatic classifier with a high probability of detection and a low false alarm rate that can classify vocalizations from a variety of cetacean species.

OBJECTIVES

In this research, we wish to apply a unique automatic classifier developed by the PI that uses perceptual signal features – features similar to those employed by the human auditory system – to classify cetacean species vocalizations and reject anthropogenic false alarms. This *aural* classifier has been successfully used to distinguish between active-sonar echoes from man-made (i.e. metallic) structures and naturally occurring clutter sources [1, 2] and performs as well or better than expert sonar operators [3]. Many of the features were inspired by research directed at discriminating the timbre of different musical instruments – a passive classification problem – which suggests it should be able to classify marine mammal vocalizations since these calls possess many of the acoustic attributes of music.

APPROACH

The research is part of a PhD program undertaken by Ms. Carolyn Binder under the supervision of Dr. Paul C. Hines. The postgraduate program is being conducted in the Oceanography department at Dalhousie University where Dr. Hines is an adjunct professor and at Defence R&D Canada–Atlantic where Dr. Hines is Principal Scientist/Underwater Sensing and Ms. Binder is a Research Assistant. In this project we examine anthropogenic transients and vocalizations from four[1] cetacean species – the sperm whale, northern right whale, the bowhead whale and the humpback whale. These species were chosen for the following reasons:

[1] Vocalization data from other cetacean species may be tested with the classifier as well, if time permits. For example, Minke whale vocalizations have recently been made available on the Mobysound website as the focal topic for the 5th International Workshop on Detection, Classification, Localization, and Density Estimation of Marine Mammals using Passive Acoustics. Including data sets such as this provide comparative a performance measures against other classifiers and tests the robustness of the classifier.

Report Documentation Page	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2012	**N/A**	**-**

4. TITLE AND SUBTITLE	
Automatic Classification of Cetacean Vocalizations Using an Aural Classifier	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Defence R&D CanadaAtlantic PO Box 1012 Dartmouth NS, Canada, B2Y3Z7La Jolla, CA 92093-0205La Jolla, CA 92093-0205**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**SAR**	**8**	

- All are present in US and Canadian waters;
- Sperm whale clicks are often confused with false alarms from impulsive anthropogenic transients and hydrophone self-noise (RF crackle, sensor knocks and bumps);
- The North Atlantic right whale is critically endangered (estimates of a few hundred remaining);
- The bowhead and the humpback have proven particularly difficult to discriminate automatically because the duration and bandwidth of vocalizations from the two species are similar.

The marine mammal vocalizations being used in the project have been obtained from several sources [5, 6]: sperm whale clicks were recorded using an SSQ57B broadband sonobuoy data files deployed from DRDC's research ship, CFAV QUEST; northern right whale vocalizations were recorded by DRDC Atlantic using a variety of sonobuoy types deployed from a Canadian Forces CP140 Maritime Patrol Aircraft; the bowhead and humpback vocalizations were obtained from the MobySound database.

The primary objective is to quantify the ability of the aural classifier to discriminate the four cetacean species from one another and from anthropogenic transients. The area A_z under the Receiver-Operating Characteristic (ROC) curve, will be used as the primary measure of performance. An additional technique to measure performance will be to examine the decisions surfaces generated by the classifier to see how well vocalizations from the species separate from one another and from the decision boundaries, and to determine what the error rates (mis-classifications) are.

A secondary objective is to examine how *robust* the classifier is. That is to say, is it likely to be useful on other vocalization data from these species collected under different environmental conditions. To examine this, discriminant analysis (DA) [7] will be used to rank the aural features in terms of their ability to separate the vocalizations between species. A subset of the most highly ranked features will be tested for robustness. To do this, a propagation experiment was conducted on board CFAV QUEST using some of the vocalizations. This experiment (facilitated through *in kind* contribution from DRDC) will be describd in the following section.

WORK COMPLETED

Primary Objective: Vocalizations from the four cetacean species mentioned previously (i.e., bowhead, humpback, North Atlantic right, and sperm whales) were used to test the classifier. A band-limited energy detector was used to process the baleen (humback, bowhead, and right whale) vocalizations and an exponential average-energy detector was used to detect the odontocete (sperm whale) clicks. The detectors were configured to allow as many detections as possible to ensure inclusion of relatively low SNR signals. Each detected vocalization was confirmed both visually (i.e. spectrogram) and aurally, and then each vocalization was placed in its own *.wav* format file with surrounding noise context. The data set consisted of 259 bowhead, 456 humpback, 142 right whale, and 178 sperm whale vocalizations – a total of 1035 signals.

The classification process begins with calculating the aural features. To do this, an auditory model is applied to each vocalization, to first obtain a perceptual representation of each signal (for more details

see reference [1]). After applying the auditory model, the dataset is divided so that half of the data in each class are used to train the classifier and other half to test it. The classifier is *trained* with vocalizations for which the classifier is provided the class label; the effectiveness of the classifier is then *tested* by imposing the assumptions of the classifier model (determined from the training set) on a dataset for which the classifier has no direct knowledge of the class label. Thus, the remaining steps are carried out using the training subset and the results are then applied to the data in the testing subset.

It is inevitable that some of the perceptual features will be more useful for discriminating between classes than others; a subset of features that best discriminate between classes can be selected, using discriminant analysis. The dimensionality of the feature space is further reduced to allow for convenient graphical representation of the results. In the reduced space, a relatively simple classifier is applied that fits a Gaussian probability density function to each class. A classification decision is made based on the largest likelihood probability of belonging to a particular class.

Secondary Objective: A CFAV QUEST trial in the spring of 2012 provided an opportunity to collect data for testing the robustness of the aural features with respect to underwater sound propagation. To investigate the impacts of propagation on aural classification, classification results of relatively high SNR ratio bowhead and humpback vocalizations can be compared to classification results obtained after the vocalizations were re-transmitted underwater over ranges of 2 to 10 km. To gain additional insight into the propagation effects, synthetic bowhead and humpback vocalizations were also transmitted. The synthetic signals were designed to have similar mean and variance values to the cetacean calls for three of the aural features found to be important to bowhead/humpback discrimination.

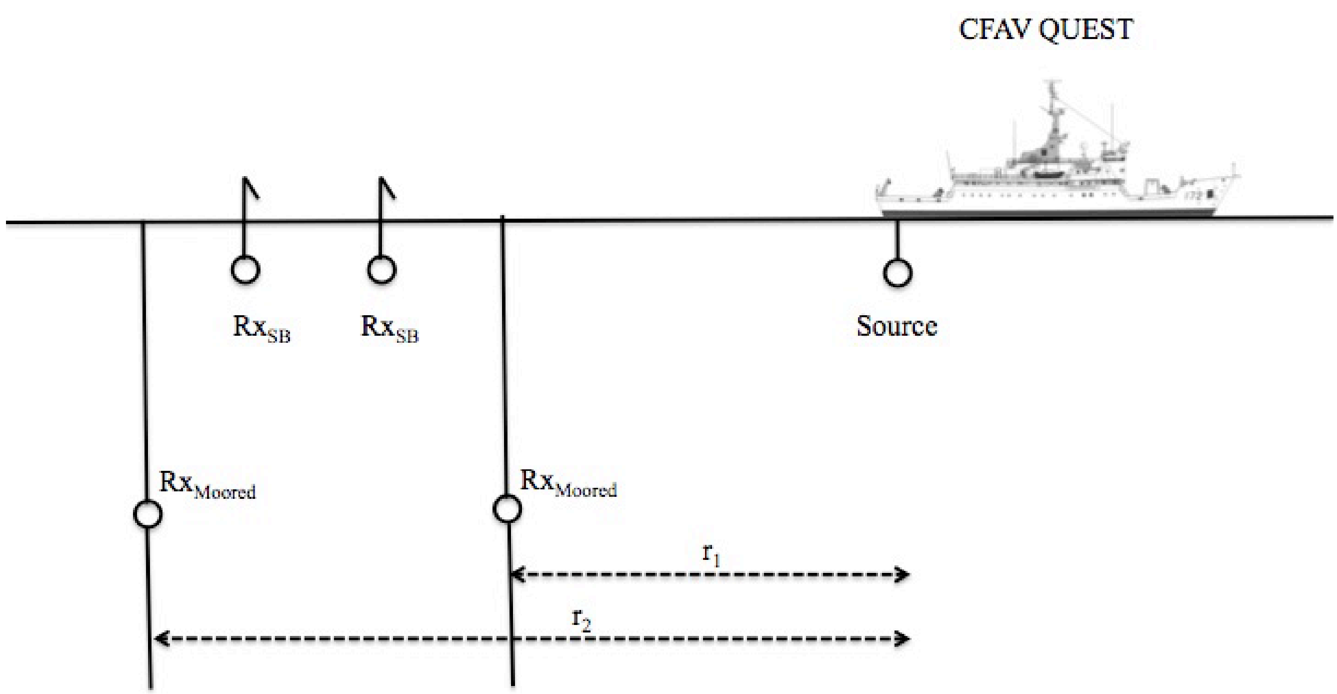

Figure 1. Experimental setup for propagation experiment. Rx_{Moored} refers to the moored recorders and Rx_{SB} refers to free-floating sonobuoy recorders. The distances between the ship and moored recorders (r_1 and r_2) ranged between 2 and 10 km.

The signals (155 of each type) were transmitted from a projector deployed from the quarterdeck of QUEST, as the ship drifted, and received on moored recorders 2-10 km away from the ship. Free-floating sonobuoy recorders with GPS locators were also used for recording the signals. The experiment was repeated three times, each on a different day (May 28, May 29 and June 2, 2012) and at a different location, so as to capture various propagation conditions. Sufficient data were obtained to start analyzing the effects of propagation on the perceptual features used by the aural classifier. Analysis of this dataset is currently being undertaken and includes examining changes to the general aural classification results, as well as examining changes to individual perceptual features to identify those features that may be robust to propagation effects.

RESULTS

The principal metric used to evaluate classifier performance is the Receiver-Operating Characteristic (ROC) curve. The ROC curve plots the probability of detecting a true positive (a correct classification) vs. the probability of detecting a false positive (an incorrect classification, sometimes referred to as a false alarm). The curve is used in a variety of disciplines where classification statistics are studied. For example, in medical diagnosis it might be used to study the success rate of detecting cancer, in which case a false positive might correspond to a benign growth being misdiagnosed as malignant. In the data presented here, a false positive would be mis-classifying one marine mammal species for another. One of the most useful (and concise) metrics one can extract from the ROC curve is the area under the curve, A_z. The greater A_z, the better the classifier; a value of $A_z = 1$ indicates ideal performance and a diagonal line ($A_z = 0.5$) represents chance performance. A single ROC curve cannot be used to evaluate classifier performance when more than two classes are considered (eg. multiclass classification of several marine mammal species). In this case performance is quantified by computing A_z for all (i,j) pairs of all c classes, using the M-measure [8]:

$$M = \frac{2}{c(c-1)} \sum_{i<j} A_z(i,j).$$

The left panel of Figure 2 shows the probability density functions (pdf) obtained by incorporating all baleen (humback, bowhead, and right whale) vocalizations into a single class and classifying against the odontocete (sperm whale). The impulsive clicks of the sperm whale are easily discriminated from the much longer duration moans of the baleen whales. Sweeping the decision boundary across the horizontal axis generates a nearly ideal ($A_z > 0.99$) ROC curve (not shown). The right panel of Figure 2 shows the normalized discriminant rank of the features used to separate the baleen and sperm whales. The names of the features are contained in Table I.

The left panel of Figure 3 shows a plot of the decision region obtained for the much more challenging case of discriminating the aurally complex vocalizations of the three baleen species. In this case, two discriminant axes are required to successfully separate the three species. If a data point is on the corresponding background colour (eg. red on pink, blue on blue, black on grey), the classifier has correctly identified it. Conversely, if data is on a different background colour, the classifier has incorrectly identified it as being from one of the other two species. The curves separating the regions define the decision boundaries. The decision surface shown in the figure corresponds to M-measures of $M = 0.98$ and $M = 0.96$ for training and testing, respectively, indicative of excellent performance. It is worth noting that projection onto a single DA axis would result in considerable overlap (and

therefore poor separation) of the humpback and bowhead species (horizontal axis) or right whale with both other species (vertical axis). The right panel of Figure 3 shows the normalized discriminant rank of the features used to separate the three baleen species shown in the left hand side of the figure. The names of the features in descending rank are contained in Table I. Since one can't generate a ROC curve for a multi-class classification, a *confusion matrix* for the pair-wise A_z values is given in Table II.

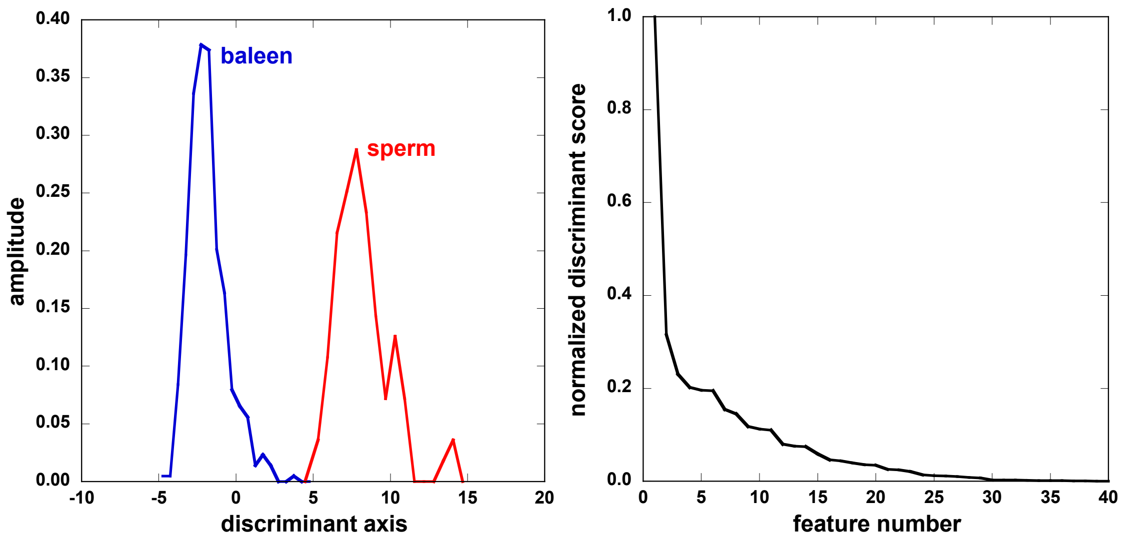

Figure 2.(Left panel) Testing DA decision region for classifying baleen and sperm whales using all non-redundant features. (Right panel) Normalized discriminant rank of the features used to separate the baleen and sperm whales shown in the left hand side of the figure.

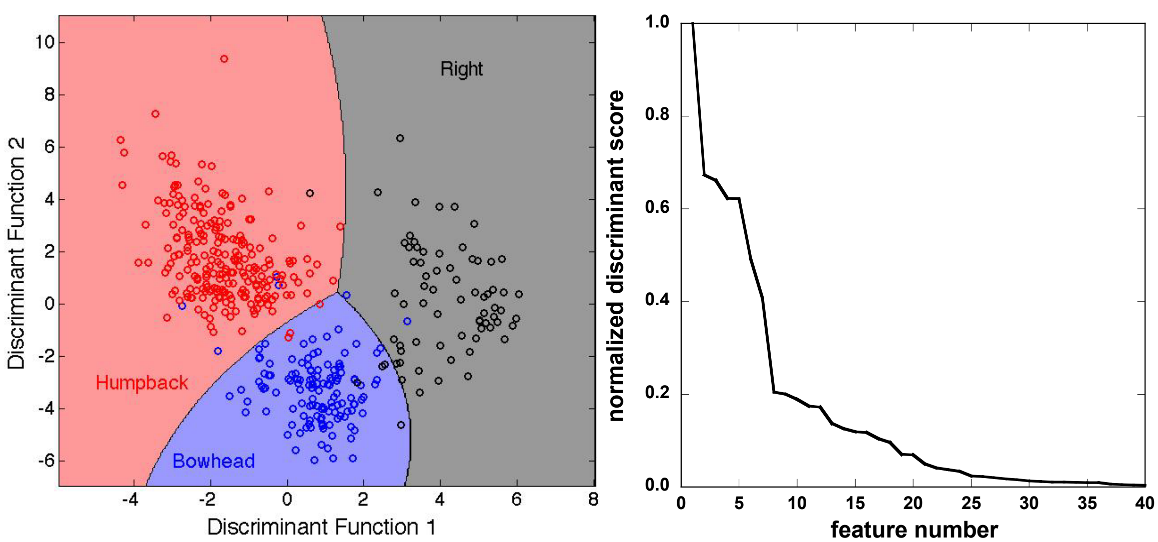

Figure 3.(Left panel) Training DA decision region for the three baleen species. (Right panel) Normalized discriminant rank of the features used to separate the three baleen species shown in the left hand side of the figure.

Table I: Top 10 features listed in rank order for the data of Figures 2 (left column) and Figure 3 (right column). A rank of 1 refers to the most important discriminating feature.

Rank	Features (Baleen vs. Sperm)	Features (Baleen)
1	Loudness Centroid	Peak loudness value
2	Frequency of global maximum sub-band attack slope	Global maximum sub-band attack time
3	Frequency of global minimum sub-band decay slope	Pre-attack psychoacoustic maxima-to-spectral-bins ratio
4	Frequency of local maximum sub-band attack slope	Mean sub-band correlation
5	Psychoacoustic maxima-to-spectral-bins ratio	Psychoacoustic maxima-to-spectral-bins ratio
6	Frequency of local minimum sub-band decay slope	Local maximum sub-band attack time
7	Frequency of local maximum sub-ban decay slope	Pre-attack integrated loudness
8	Frequency of local minimum sub-band attack slope	Local mean sub-band decay slope
9	Pre-attack psychoacoustic maxima-to-spectral-bins ratio	Local mean sub-band attack time
10	Global maximum sub-band attack time	Global mean sub-band decay slope

Table II: Confusion matrices showing pair-wise A_z values for testing and training data obtained from the aural classifier. The asterisk shows values that appear ideal due to rounding.

Train	**Humpback**	**Right**	Test	**Humpback**	**Right**
Bowhead	0.99	1.00*	**Bowhead**	0.88	1.00*
Humpback		1.00*	**Humpback**		1.00

IMPACT/APPLICATIONS

Detection and classification of cetaceans has become critically important to the US Navy due to an ever increasing requirement for environmental stewardship. Passive acoustics continues to be the best method to carry out this task but current techniques provide only a partial solution; most detectors are either too specialized (i.e., species-specific) leading to many missed detections, or are too general, leading to unacceptably high false alarm rates. Furthermore, future military platforms will have to support smaller complements and deal with ever-increasing data throughput, so that automation of on-board systems is essential. In addition, the technique is well suited to autonomous systems since a much smaller bandwidth is needed to transmit a classification result than to transmit raw acoustic data. The success of the aural classifier in discriminating cetacean vocalizations suggests that it could be applied to other passive acoustic classification problems which currently employ human audition. This would be particularly useful if expert listeners aren't available –such as diagnosing heart murmurs in remote communities that lack a cardiologist, or as part of the triage process in a hospital emergency department. Alternatively, the aural classifier is ideally suited when the sheer volume of data makes human audition untenable – such as classifying ocean acoustic data for species population monitoring.

Finally, testing the classifier on passive marine mammal vocalizations is also a first step to testing the algorithm on passive transients generated by submarines to examine its potential for passive detection and classification of submarines.

RELATED PROJECTS

This research will benefit from DRDC Atlantic's SUBTRACTION Applied Research Project in which DRDC's aural classification algorithms (including the marine mammal classification algorithm) will be integrated into DRDC's System Test Bed (STB). The STB is used to evaluate sonar algorithms in a military context. Some of the insights to be gained will be: whether the aural classifier can reduce false alarms from marine mammals; does the classifier reduces operator workload required by environmental considerations (the so-called green navy) to enable greater concentration on potential targets; is the aural classifier easily integarted into a navy platform. This research also benefits substantially from a recently completed project at DRDC [6] during which anthropogenic transients and cetacean vocalization data were compiled, extracted into *.wav* files, and manually classified with assistance from expert listeners.

REFERENCES

[1] Victor W. Young and Paul C. Hines, "Perception-based automatic classification of impulsive-source active sonar echoes," J. Acoust. Soc. Am., 122 (3) pp. 1502-1517, (2007).

[2] Paul C. Hines, Victor W. Young, and Jeff Scrutton, "Aural classification of coherent-source active sonar echoes," Proceedings: International Symposium on Underwater Reverberation and Clutter, Lerici, Italy, eds. P.L. Nielsen, C.H. Harrison, and J.-C. LeGac, pp. 293-301, (2008).

[3] Nancy Allen, Paul C. Hines, and Victor W. Young "Performances of human listeners and an automatic aural classifier in discriminating between sonar target echoes and clutter," J. Acoust. Soc. Am., 130 (3), pp. 1287-1298, 2011.

[4] Carolyn M. Binder, Paul C. Hines, and Joey Hood, "Passive Classification of Marine Mammal Vocalizations Using an Automatic Aural Classifier," 5th International Conference on Detection, Classification, Localization, and Density Estimation of Marine Mammals Using Passive Acoustics, Mt. Hood, Oregon, August, 2011.

[5] S. Heimlich, D. Mellinger, H. Klinck, "MobySound.org," http://www.mobysound.org/index.html (2007).

[6] Joe Hood and Derek Burnett, "Compilation of Marine Mammal Passive Transients for Aural Classification," DRDC Atlantic CR 2008-287, 93 pages, (2009).

[7] R.O. Duda, P.E. Hart, and D.G. Stork , Pattern Classification, (2nd ed, Wiley-Interscience, 2001).

[8] David J. Hand and Robert J. Till, "A simple generalisation of the area under the ROC curve for multiple class classification problems," Machine Learning, vol. 45, pp. 171-186, 2001.

PUBLICATIONS

[1] Paul C. Hines, Carolyn M. Ward, and Joey Hood "A Hearing-Inspired Classifier To Discriminate Inter-Species Marine Mammal Vocalizations, Proceedings: 18th International Congress on Sound and Vibration 2011, Rio de Janeiro, Brazil, July, 2011. [published, refereed]

[2] Carolyn M. Ward, Paul C. Hines, and Joey Hood "Passive Classification of Marine Mammal Vocalizations Using an Automatic Aural Classifier", Proceedings: 5th International Workshop on Detection, Classification, Localization, and Density Estimation of Marine Mammals using Passive Acoustics, Mt. Hood, Oregon, August, 2011. [published]

[3] Carolyn M. Binder, "Using an Automatic Aural Classifier to Discriminate Cetacean Vocalizations," M.Sc. Thesis, Dalhousie University, Halifax NS (2012). [published]

[4] Carolyn M. Binder and Paul C. Hines, "Applying Automatic Aural Classification to Cetacean Vocalizations," Proceedings: 11th European Conference on Underwater Acoustics, Edinburgh, UK, pp. 323-330 (2012). [published]

HONORS/AWARDS/PRIZES

Carolyn Binder (Defence R&D Canada – Atlantic / Dalhousie Unversity), "European Conference on Underwater Acoustics 2012 Student Presentation Prize," sponsered by the UK Institute of Acoustics.